

Joint effects of nine polychlorinated biphenyl (PCB) congeners on breast cancer risk

Theodore R Holford,^a Tongzhang Zheng,^a Susan T Mayne,^a Shelia H Zahm,^b John D Tessari^c and Peter Boyle^d

Background	Polychlorinated biphenyls (PCB) have been a major environmental health concern because of their wide distribution and persistence in the environment. Estimating joint effects of all congeners in a single analysis is complicated by correlation among exposure levels, and the resulting collinearity makes the results difficult to interpret.
Methods	Patients with breast-related surgery at Yale-New Haven Hospital were interviewed using a standardized questionnaire, and breast adipose tissue samples were analysed for nine PCB congeners (74, 118, 138, 153, 156, 170, 180, 183, 187). The study recruited 490 women (304 cases and 186 controls) between 1994 and 1997. Logistic ridge regression was used to analyse the instability caused by collinearity.
Results	Although total PCB did not appear to be associated with breast cancer risk, significant differences in effect were observed among the nine congeners. Logistic ridge regression demonstrated a protective effect on breast cancer risk for a potentially anti-oestrogenic and dioxin-like congener, 156, while two phenobarbital, CYP1A and CYP2B inducers had an adverse effect, 180 and 183. This analysis also suggested that a protective effect for another phenobarbital congener, 153, was largely explained by instability caused by collinearity.
Conclusions	These results indicate that studies of PCB congeners and health require an in-depth statistical analysis in order to better understand the complex issues related to their collinearity.
Keywords	Breast cancer, environmental epidemiology, PCB, logistic regression, ridge regression
Accepted	16 May 2000

Polychlorinated biphenyls (PCB) have been a major environmental health concern because of their wide distribution and persistence in the environment. While commonly perceived as a single entity, they actually represent a large class of chemicals that includes 209 different congeners.¹ Depending on the degree of chlorination, a PCB may be a mobile oil, a viscous liquid or a solid. The invaluable industrial uses of these compounds arise from their chemical stability, their dielectric properties, and their inflammability. Unfortunately, careless disposal over decades has resulted in their wide-spread distribution in the ecosystem, where these same stability properties have resulted in their persistence as a potential environmental health hazard.²

Potential mechanisms by which PCB could affect risk of diseases like breast cancer include various types of hormonal activities that have been observed for specific congeners. Some exhibit oestrogenic behaviour,^{3,4} thus possessing the theoretical potential for increasing breast cancer risk. Indeed, Dewailly *et al.*⁵ suggest that accumulation of such organochlorines may increase risk for hormone-responsive breast cancers. On the other hand, some congeners behave as anti-oestrogens, which could in principle counteract the putative harmful effects of oestrogen and result in an ameliorative effect on breast cancer risk, not unlike the possible protection offered by tamoxifen.⁶ Wolff and Toniolo⁷ have suggested that the common practice of analysing epidemiological studies by considering total PCB as a putative risk factor could be missing important nuances in exposure. If the putative effects of PCB arose because of oestrogenic activity, then it would make better sense to analyse these exposure groups separately. Other factors that could be considered in such groupings include persistence, dioxin-like activity, phenobarbital induction, potential for inducing cytochrome P450 enzymes,^{4,7} and degree of chlorination.⁸ An analysis that included these factors would reflect the underlying biology thought to be relevant in the aetiology of breast

^a Department of Epidemiology and Public Health, Yale University School of Medicine, New Haven, CT, USA.

^b National Cancer Institute, Bethesda, MD, USA.

^c Department of Environmental Health, Colorado State University, Fort Collins, CO, USA.

^d European Institute of Oncology, Milan, Italy.

Reprint requests: Dr Theodore R Holford, Department of Epidemiology and Public Health, Yale University School of Medicine, 60 College Street, PO Box 208034, New Haven, CT 06520–8034, USA.

cancer. However, this approach could still miss potentially important combinations of congener exposure if the underlying mechanism, and thus the appropriate grouping of congeners, was unknown.

Commercial PCB are marketed by the proportion of chlorine by weight, so that a variety of congeners are usually present in a particular product.¹ Hence, exposure to one congener is likely to be correlated with exposure to others that happened to be in the same product. The resulting high correlation among the exposure levels for individual congeners can result in collinearity, which confounds the estimates of effect on breast cancer risk. For example, if the exposure to an oestrogenic congener was highly correlated with an anti-oestrogenic congener, then the effect of either one on breast cancer risk might not be apparent because when one was high, the other would also be high and their combined effect could tend to cancel each other out. In this hypothetical case, a better analysis might be to include the difference in exposure levels for these two congeners as a covariate in the analysis instead of their total.

We consider below the joint effect of nine PCB congeners measured in breast adipose tissue on the risk of breast cancer. These data are from a case-control study in Connecticut, and the effect of total PCB exposure along with the effects of individual congeners taken one-at-a-time was considered in an earlier manuscript.⁹ Because the exposure levels for individual congeners were highly correlated, collinearity could have a profound effect on the estimated associations. The statistical techniques used to assist in the interpretation of these results included principle components analysis and ridge regression.

Methods

Study population

A total of 490 women (304 cases and 186 controls) were recruited into the study between 1994 and 1997 after obtaining consent from the patients themselves and their physicians.⁹ Study subjects were patients who had breast-related surgery at Yale-New Haven Hospital (YNHH), Connecticut, USA. Cases were histologically confirmed, incident primary breast cancer patients (ICD-O, 174.0–174.9) aged 40–79, and controls were patients with histologically confirmed incident benign breast disease (including a diagnosis of normal breast tissue, but excluding atypical hyperplasia). Neither cases nor controls had a previous diagnosis of cancer, with the possible exception of non-melanoma skin cancers. The participation rates for the study were 79% for the cases and 74% for the controls among those having at least 0.4 g of adipose tissue for organochlorine compound analyses.

Breast adipose tissue and chemical analysis

Breast adipose tissue not needed for diagnostic purposes was collected and placed into a glass bottle on ice, coded and frozen within 30 minutes after biopsy. Samples were stored at –84°C until being packed in blue ice and mailed in batches to the study laboratory at Colorado State University, where they were stored frozen until analysis. Laboratory personnel did not know the case-control status of samples.

Nine PCB congeners that are known to be abundant in human adipose tissue were measured in this study, International Union of Pure and Applied Chemists (IUPAC) congener numbers: 74,

118, 138, 153, 156, 170, 180, 183, and 187. These congeners were found to have higher concentrations than others in preliminary analyses of some study samples, and thus they were likely to provide the greatest precision of measurement in the chemical analysis. Stellman *et al.*¹⁰ report the relative abundance of 14 PCB in Long Island, NY, and the reported means for our nine congeners represent 87% of their total mean concentration. Earlier data reported by McFarland and Clark¹ provide per cent of total PCB for what was known about all 209 congeners, and our nine congeners comprise 59% of the total found in human fat, compared to 66% for Stellman *et al.*'s 14.

The laboratory method for analysing PCB congeners in breast adipose tissue has been described elsewhere.¹¹ The method involves extraction of the compounds of interest in hexane, separation of the organochlorine pesticides from the PCB and purification of the sample using Florisil® chromatography, and identification and quantification of the compounds using gas chromatography. The quantitation limits (ppb) of this method were: 7.5 for 74; 12.5 for 118; 10.0 for 138; 25.0 for 153; 25.0 for 156; 20.0 for 170; 12.5 for 180; 10.0 for 183; and 12.5 for 187.

Strict quality control/quality assessment procedures were followed throughout sample analysis, including method spikes, reagent blanks, and the establishment of quality control windows. Quality control (QC) spike recoveries for the PCB throughout the sample analyses ranged from 83% to 108% with a coefficient of variation (CV) that ranged from 12% to 24%. The QC spike recoveries for the PCB isomers during the sample analyses ranged from 82% to 96% with a CV of 10–25%. Adipose tissue levels of PCB congeners were reported as parts per billion (ppb), which is equivalent to nanograms of PCB congener per gram of lipid. The amount of lipid in the sample was quantified gravimetrically.

Statistical methods

The data analyses were based on lipid-adjusted adipose tissue levels of PCB using a linear logistic regression model as implemented in PROC GENMOD in SAS.¹² Covariates included in the final model were age, body mass index, lifetime months of lactation, age at menarche, age at first full pregnancy (<25 and ≥25), number of live births (none, <3, and ≥3), dietary fat intake, household income (<\$8750, \$8750–14 284, 14 396–24 999, ≥25 000, and unknown), and fat levels of DDE (<435.2, 435.2–784.3, 784.4–1437.3, ≥1437.4). DDE was included as a covariate because earlier studies suggested that it might be associated with breast cancer risk.

Of primary interest in this analysis was the joint effects of individual PCB congeners on risk of breast cancer, and whether the effect of each congener was the same, which was tested using a linear contrast. If these results suggested that the magnitude of effect on breast cancer risk was different among the congeners then it would not make sense to evaluate total PCB exposure, but to investigate the joint effects of each congener. Regression diagnostics were used to determine whether the results were sensitive to one or more influential observations, but the overall conclusions were found to be stable. Bootstrap methods were used to estimate bias in the estimates of risk, as well as providing alternative estimates of standard errors.¹³ While the resulting standard errors were slightly greater, the conclusions were essentially unchanged, so these results are not presented.

If individual congener levels are highly correlated, as was the case, then collinearity must be considered. It is well known that when the collinearity is extreme, the numerical accuracy of the results can be affected. While we did not find evidence of such inaccuracies in this analysis, the correlations were high enough to produce profound effects on the estimated associations with breast cancer risk. This was addressed in the analysis by (a) variable reduction, (b) principle components, and (c) ridge regression, which are described below.

Variable reduction

The simplest way of dealing with collinearity is to drop redundant variables from the regression model. This approach is predicated on the idea that only a subset of variables is needed in the regression model. If, on the other hand, we have two chemicals that have highly correlated exposure levels, and they both affect the risk of disease, it may not be possible to accurately estimate their separate effects on risk. Dropping one of the chemicals from a model may not change the accuracy of the prediction, but it forces that regression coefficient to be zero, thus providing a biased estimate of effect for both chemicals.

Principle components

An alternative to dropping variables from the analysis is to consider linear combinations that make substantive scientific sense. One of the simplest and most commonly used summaries is to analyse the effect of total PCB. Principle components analysis offers another approach for identifying components of variation among the factors of interest, predicated on the idea that there are common unmeasured factors giving rise to an observed joint distribution of exposures. In order to understand better the nature of the effects for individual congeners, principle components analysis was used to create factors that were independent of each other. Using PROC PRINCOMP in SAS we estimated the eigenvectors, which provided loading scores that gave rise to new variables to be included in the linear logistic model.

Ridge regression

Yet another approach for dealing with instability of parameter estimates in the presence of collinearity is ridge regression.^{14,15} Extensions of this idea to binary outcomes and/or the generalized linear model are provided by Schaefer^{16,17} and Segerstedt.¹⁸ To employ this method we first normalized the exposure measures for each congener by subtracting the mean for the study population and dividing the result by the standard deviation for the congener. Maximum likelihood estimates of the logistic regression parameters for each congener were obtained

using PROC GENMOD in SAS, first adjusting for age alone, and then the remaining covariates. These results were then used to obtain ridge regression estimators for a particular ridge coefficient, k ($k \geq 0$), by implementing the formulae described in the Appendix using PROC IML of SAS.

A fundamental issue in ridge regression is the selection of the ridge coefficient, k . When $k = 0$, the result is the usual maximum likelihood estimator, and as k becomes large the ridge estimators eventually go to 0, although before reaching the limit they can change sign. Parameter estimates that are heavily influenced by collinearity tend to change rapidly for small values of k , and become more stable as k increases. This phenomenon can be observed by creating a ridge trace, which plots the ridge estimators against k . The reported ridge estimator of a regression parameter uses a small value of k in the range in which the ridge trace has been stabilized. While the resulting estimator is no longer a maximum likelihood estimate, and is thus biased, it will generally have reduced variance. Our objective is to find estimates that have a reduced mean squared error, which is the sum of the variance and the square of the bias. The results from this approach are not unique in the sense that a certain amount of judgement is necessary when interpreting the ridge trace, but the fact that the choice of k is in the range where the trace is stable means that some variation in this coefficient will not substantially change estimates of effect.

Results

The distribution of some of the typical risk factors for breast cancer are shown in Table 1 of Zheng *et al.*,⁹ and Table 1 of Zheng *et al.*,¹⁹ so this will not be repeated here. The only differences among cases and controls that were particularly noteworthy were that the cases were 3.7 years older on average ($P < 0.01$), and the parous cases lactated 1.2 fewer months ($P = 0.05$).

One individual had a value for IUPAC congener number 183 that was an outlier with an exposure 4.6 standard deviations over the mean on the log scale (Bonferroni $P = 0.0020$), and this individual was excluded from further analyses that are reported here. Analyses were also conducted with this observation included, and the results were essentially unchanged. Figure 1 shows a scatter plot matrix for each of the nine congeners measures in this study, and Table 1 provides summary statistics and correlation coefficients. Clearly the exposures to specific congeners are highly correlated, thus giving rise to a

Table 1 Summary statistics and correlation coefficients for polychlorinated biphenyl (PCB) congeners

	74	118	138	153	156	170	180	183	187
118	0.68								
138	0.67	0.82							
153	0.58	0.66	0.86						
156	0.40	0.52	0.58	0.54					
170	0.52	0.59	0.80	0.78	0.71				
180	0.45	0.43	0.67	0.87	0.47	0.79			
183	0.46	0.53	0.70	0.77	0.65	0.77	0.69		
187	0.53	0.50	0.68	0.84	0.47	0.74	0.83	0.75	
Mean	36.16	52.39	95.04	143.29	24.80	37.73	110.32	14.52	34.58
SD	30.11	52.28	64.61	76.50	16.15	22.24	58.09	8.98	22.21

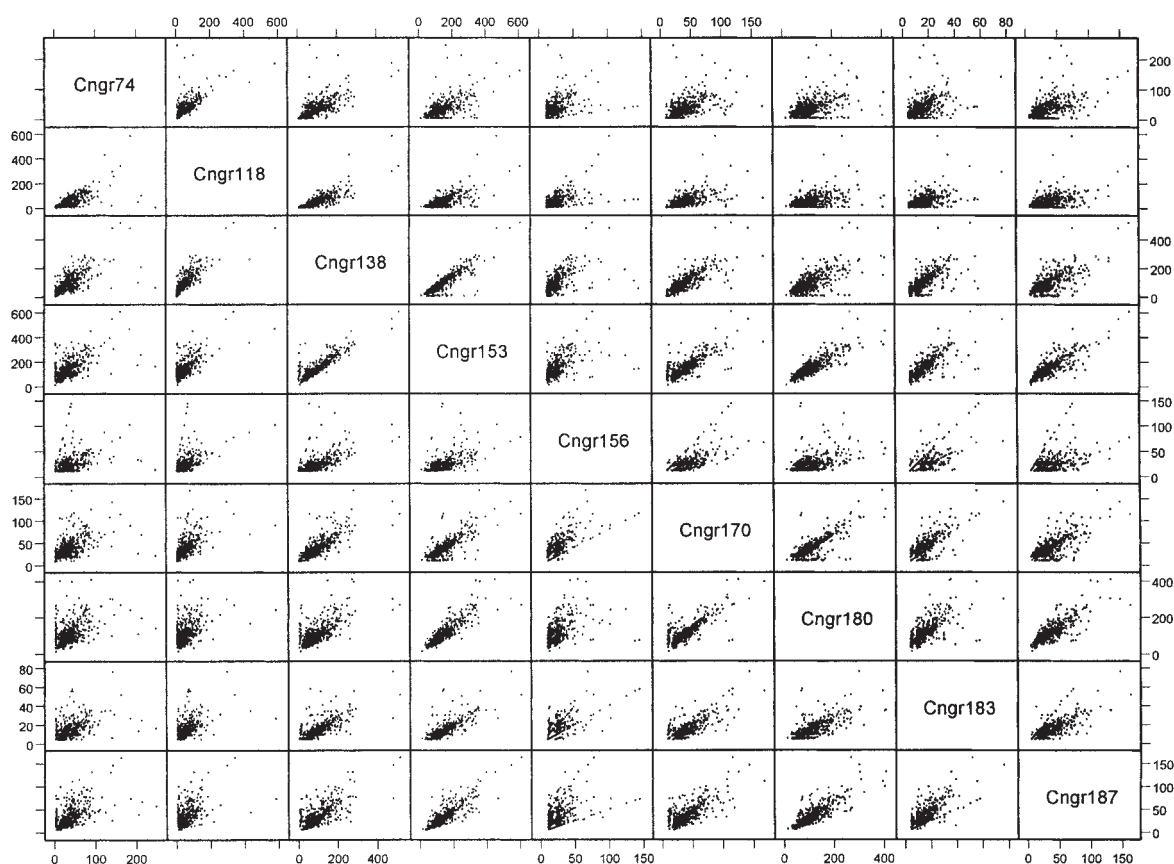


Figure 1 Scatter plot matrix for individual PCB congeners

potential problem with interpreting the results due to collinearity. However, they are not so extreme that the numerical tolerance limits for computational accuracy have been exceeded.

The results of an analysis of total PCB exposure suggested a neutral effect odds ratio, (OR) = 0.97 (95% CI : 0.90–1.05) for a 100-ppb change in exposure that did not achieve nominal

statistical significance ($P = 0.498$). Table 2 presents estimates of the relative risk for breast cancer associated with a 10-ppb change in exposure to individual congeners, obtained by fitting a linear logistic model to the data. Notice that some congeners are positively associated with breast cancer risk while others are negative, and the global test of all congeners entered simultaneously was

Table 2 Estimates of the relative risk for breast cancer associated with a 10-ppb change in exposure to individual congeners by type of model

	Standard coefficient ^a	All congeners ^a	Risk-associated congeners ^{a,b}	Ridge ^c -age only	Ridge ^c -covariates ^a
74	–0.21	0.93 (0.84–1.04)		0.98 (0.92–1.04)	0.96 (0.90–1.03)
118	0.19	1.04 (0.96–1.12)		0.99 (0.96–1.03)	1.00 (0.97–1.04)
138	0.28	1.04 (0.94–1.16)		1.00 (0.97–1.03)	1.00 (0.97–1.02)
153	–1.04	0.87 (0.78–0.98)	0.93 (0.86–0.99)	0.98 (0.96–1.00)	0.98 (0.96–1.01)
156	–0.37	0.79 (0.64–0.99)	0.78 (0.65–0.93)	0.87 (0.77–0.98)	0.87 (0.78–0.99)
170	–0.36	0.85 (0.65–1.11)		1.00 (0.91–1.09)	0.99 (0.90–1.08)
180	0.74	1.14 (1.00–1.29)	1.09 (0.99–1.19)	1.04 (1.01–1.07)	1.02 (0.99–1.05)
183	0.54	1.82 (1.12–2.98)	1.74 (1.12–2.71)	1.22 (0.98–1.54)	1.23 (0.98–1.54)
187	0.24	1.11 (0.90–1.37)		1.04 (0.95–1.14)	1.05 (0.96–1.15)
LR χ^2		18.09 (d.f. = 9)	14.30 (d.f. = 4)		

^a Adjusted for age, and months of lactation, body mass index, fat consumption, age at menarche, number of live births (none, <3, and ≥ 3), age at first pregnancy (<25 and ≥ 25), income (>\$8750, 8750–14 284, 14 396–24 999, ≥ 25 000, and unknown), fat levels of DDE (<435.2, 435.2–784.3, 784.4–1437.3, ≥ 1437.4).

^b Congeners that excluded the null odds ratio in the initial analysis.

^c Ridge regression estimate of OR for a 10-ppb change in exposure for $k = 20$.

nominally significant ($P = 0.034$). The effect of total PCB would give a reasonable estimate of each congener if their effect was identical. By comparing the likelihood ratio test for the model with total PCB with the one with individual congeners, we obtain a test of whether the effects are identical, which is rejected ($\chi^2 = 17.63$, d.f. = 8, $P = 0.024$). Hence, we have quite strong evidence that the effects of these individual congeners are not identical. To check the validity of the linearity assumption inherent in the use of this model, we categorized exposure for each congener into quartiles and tested for goodness of fit for a linear trend using mean dose for each quartile. Only one approached statistical significance (183, $P = 0.057$) which is not surprising in light of the fact that nine tests were conducted, and all of the other congeners had P -values for goodness of fit of ≥ 0.225 . Hence, we concluded that the linearity assumption adequately represented these data.

As part of our consideration of the possibility that these results are only an artifact of collinearity, we constructed a PCB breast cancer risk score for each individual by multiplying each congener level by its corresponding regression coefficient, and adding them up. The standardized regression coefficients in Table 2 estimate the change in the log relative risk for a one standard deviation change in exposure for an individual congener, and thus are an indication of the degree to which an individual congener affects the risk score. These risk scores were then categorized into quintiles based on the distribution among the controls (Table 3). The trend with this risk score is necessarily increasing because it represents the PCB contributions to the logit of the response. However, the direction of the contribution of individual PCB to this score depends on the sign of the corresponding regression coefficient, which would be negative for a 'protective' congener and positive for a 'harmful' one.

To consider further the extent to which collinearity affects our estimates of association, we first dropped the five congeners that were not significantly different from the null, and focused on just the four that appeared to be associated with breast cancer risk: 153, 156, 180 and 183. Each of the four included congeners retains its statistical significance, and the joint test of the five excluded was not significant ($\chi^2 = 3.79$, d.f. = 5, $P = 0.580$). The fact that the χ^2 statistic would not achieve statistical significance even if d.f. = 1 suggests that little is lost by only considering the subset of four congeners. Table 4 gives the eigenvectors for the principle components for the exposure to these congeners, and each component for a particular subject was determined by taking the vector product of the eigenvector and the levels of exposure. The first principle component accounts for 75% of the variance, and the elements of the

Table 4 Eigenvectors for the principle components for four congeners and the statistical significance of their association with breast cancer risk

Congener	Principle components			
	#1	#2	#3	#4
153	0.535	-0.329	0.093	-0.773
156	0.431	0.804	0.409	0.005
180	0.509	-0.475	0.391	0.602
183	0.519	0.138	-0.819	0.202
Proportion of variance explained	0.75	0.15	0.06	0.03
Test of association with breast cancer				
χ^2	0.02	4.13	3.55	5.67
P -value	0.898	0.042	0.060	0.017

eigenvector are nearly identical, so that an analysis of this component is essentially the same as the results from taking the total exposure for these congeners. Also shown in Table 4 are the results obtained when these components are included in a model predicting breast cancer risk. Only the first principle component is not significant. The remaining components involve exposure differences among congeners, and each attains at least marginal significance.

The ridge trace for an analysis based on all congeners and adjusting for age only is shown in Figure 2. The logistic regression coefficients shown in the graph reflect the contribution of normalized exposures, i.e. they represent the effect of one standard deviation change in exposure. While some coefficients do change sign, most tend to be near the null value of zero. While 153 shows the largest protective effect, its effect largely disappears when the trace stabilizes as k nears 20, at which point 156 has a stronger protective effect. Congeners 180 and 183 have the largest adverse effect on breast cancer risk. Table 2 presents the ridge estimate ($k = 20$) for each congener, adjusted for age only. The ridge trace was also completed for an analysis that adjusted for the other covariates, and the results were very similar. Covariate adjusted ridge estimates for the congener effects are shown in Table 2.

Discussion

The association of total PCB exposure with breast cancer risk in this analysis was estimated to be small and inverted, which is consistent with the results of earlier studies. However, we found overall significance when all individual congeners were included in the analysis, and we also found significant differences in the effects of the nine congeners measured. These results suggest that some congeners have a protective effect on breast cancer risk, e.g. congener 156, while others are associated with an increased risk, e.g. congeners 180 and 183.

One difficulty in interpreting these results is the fact that levels of exposure to individual congeners are correlated, and thus collinear. While very highly correlated factors can cause the numerical accuracy of regression parameters to suffer, this was not the case for these data. In fact, the likelihood ratio statistics used to test the overall significance of the individual congeners and the differences in the congener-specific effects does not depend on a matrix calculation, which is the type of computation most affected by collinearity. Instead, our

Table 3 Distribution of subjects by quintiles of polychlorinated biphenyl (PCB) related risk score, and adjusted estimates of relative risk

Quintile	Controls	Cases	OR ^a
1	36	31	1.00
2	37	41	1.45 (0.69–3.04)
3	38	49	1.71 (0.83–3.55)
4	37	58	1.61 (0.78–3.33)
5	39	124	3.57 (1.78–7.17)

^a See footnote for Table 2.

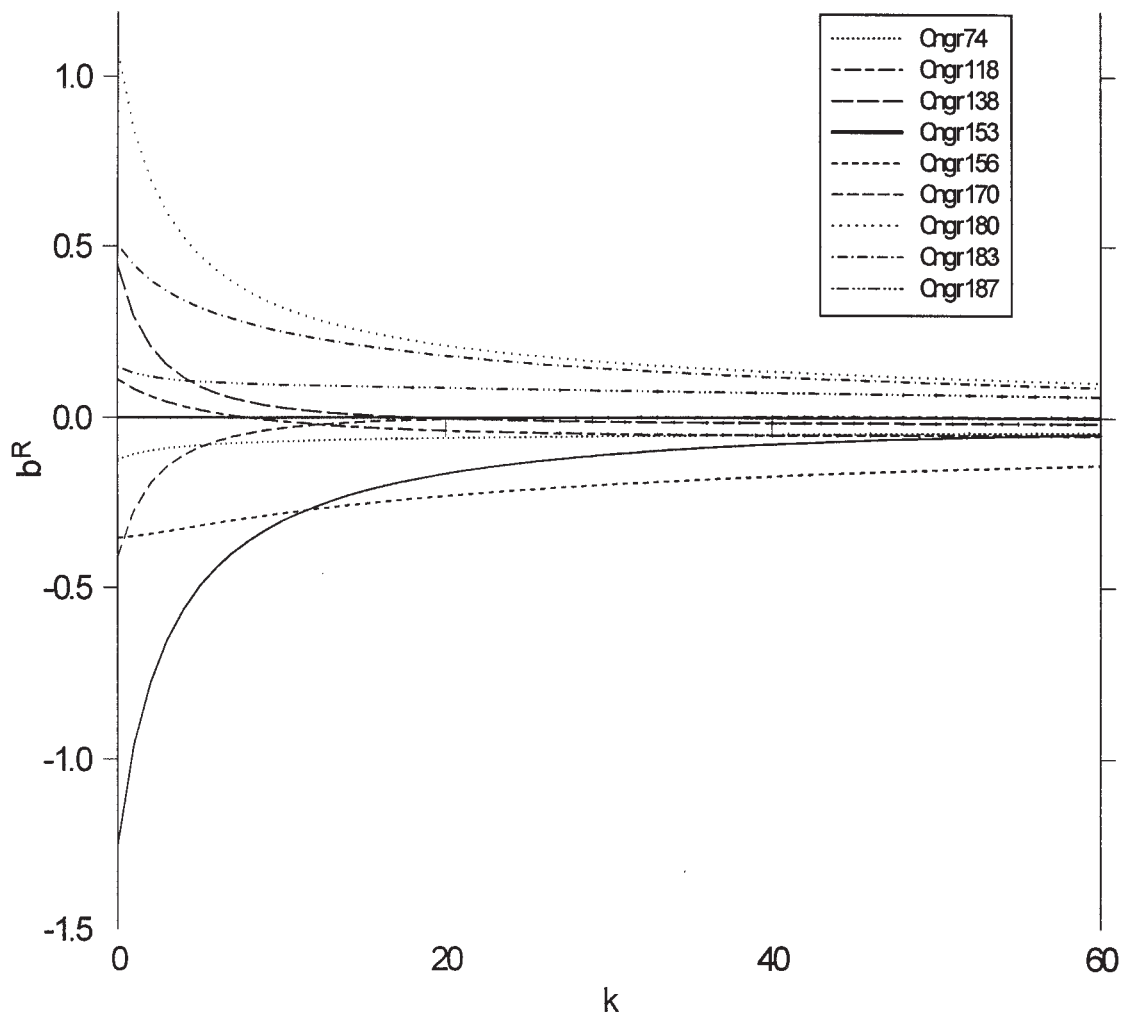


Figure 2 Ridge trace for individual congeners adjusted for age only

primary concern here is with interpretational difficulties of the effects, and the extent to which they can be attributed to specific congeners.

We have shown that our model is able to identify groups of women with very different levels of breast cancer risk. The trend in risk by the quintiles formed from the breast cancer risk score (Table 3) is apparent. In addition, the highest quintile has 3.57 times the risk of those in the first, which is a sizeable effect. Hence, it is clear that this approach is able to separate these women into categories with important differences in breast cancer risk.

The principle components analysis showed that 75% of the overall variation in the levels of exposure for congeners 153, 156, 180 and 183 was explained by the first component, which was similar to their total. Because this was not significant, our results are consistent with studies that have not found an effect due to total PCB exposure. However, we have found significant differences in the effects of the individual congeners.

Ridge regression allowed us to determine the individual congener effects that were most affected by collinearity. The effects

of 180 and 183 remain positive at the value of the ridge coefficient that stabilizes the parameters ($k = 20$). In addition, the estimated protective effect of 156 is little changed by varying the ridge coefficient. The effect of 153, on the other hand, exhibits a profound change, from a strong protective effect to one that largely disappears when the ridge coefficient stabilizes the parameter estimates.

Wolff *et al.*⁴ have proposed a grouping of the congeners as a means of interpreting their possible effects on breast cancer risk. Congener 156 is in Group 2, whose members are anti-oestrogenic and dioxin-like, and thus could interfere with the harmful effect of oestrogen to provide a protective effect on breast cancer risk. On the other hand, congeners 180 and 183 fall into Group 3, the phenobarbital, CYP1A and CYP2B inducers, and the manner in which these might affect breast cancer risk is far from clear. Some have suggested that induction of CYP1A1 gene expression could lead to a protective effect.² Yet, two epidemiological studies have found that CYP1A1 polymorphism may increase risk of breast cancer especially among some smokers.^{20,21} A third epidemiological study failed to find such

an association,²² so much remains to be learned about the aetiology of breast cancer, but at this point it is impossible to predict the manner in which a chemical might affect its risk. Our results are broadly consistent with what might be expected for both of these groups. Ridge regression suggested that the significant protective effect for 153 that was observed initially could largely be accounted for by the instability caused by collinearity. Because 153 belongs to Group 3 along with 180 and 183, one would not expect the opposite effects observed in the initial analysis. However, further study revealed that the protective effect was largely due to an artifact of collinearity. Hence, the ridge estimators of the effects of the individual congeners are consistent with what one might expect from the congener groups proposed by Wolff *et al.*⁴

While our results are broadly consistent with the categories of PCB congeners proposed by Wolff *et al.*,⁴ there still appears to be some differences in the effects of congeners within these broad groups. The use of these categories assumes a knowledge of the mode of action with respect to breast cancer risk, but this is probably premature. It is also possible that while a category might suggest a direction for an effect, the magnitudes could still differ.

The PCB are readily absorbed from the gut in mammals and very little is excreted in its parent form, so that they tend to build up in adipose tissue because of their lipid solubility.²³ The rate of biotransformation for PCB tends to decrease with the number of chlorines on the biphenyl nucleus; and it varies with the position of the chlorines, as well as the species. Dogs and rats, for instance, tend to metabolize PCB much more readily than monkeys. For example, 50% of congener 136 is eliminated by dogs and rats in one day, while monkeys require 6 days. On the other hand, congener 153 which like 136 has six chlorines, has a half-life of 8 days in dogs, but monkeys and rats are apparently not able to eliminate 50% of the administered dose over their remaining lifespan.²⁴ Wolff *et al.* found a median half-life for total PCB of 37 years among cases in the New York University Women's Health Study, but it was indeterminate among controls.²⁵ Clearly, some of these compounds remain in the body for a very long time, but much remains to be learned about their biotransformation in humans.

Unlike most of the earlier studies of the effect of PCB on breast cancer risk, we determined levels in breast adipose tissue instead of blood serum. Tessari and Archibeque-Engle²⁶ have pointed out the analytical difficulties with the chemical measures of the extremely low levels of PCB found in sera. In addition, serum levels can be very sensitive to short-term changes in diet, such as fasting. For a subset of these cases, we also measured serum levels of individual congeners and, while we found that the two are correlated (data not shown here), sera offer an imperfect surrogate for the levels found in breast adipose tissue. It is well known that even purely random errors in exposure measurements can bias estimates of association with disease risk. This could account for the absence of congener specific effects observed by Moysich *et al.*,⁸ who determined serum levels of PCB.

Our use of breast adipose tissue precluded the possibility of using general population controls in this study, requiring instead the recruitment of women without breast cancer who had undergone breast surgery. Many of these subjects had other benign breast diseases, some of which may also be related to

hormonal factors. If the effect of PCB on the breast diseases in the controls is in the same direction as that of breast cancer, then our study would be biased in the direction of underestimating the effect of PCBs on breast cancer. Alternatively, if some of the benign conditions in the controls are really precursors of cancer, the direction of the bias would again be to attenuate estimates toward the null, especially if the effect of PCB occurs early in the disease process.

A limitation of this study is the fact that only 9 of the 209 PCB congeners were measured for each subject. While these are among the most common congeners found in humans, comprising 87% of the total mean concentration found in a Long Island, NY study,¹⁰ it is still possible that we have measured compounds that are actually surrogates for others that are the real causative agents for breast cancer. We have noted the collinearity among the congeners we did measure, so it is not a large stretch to consider the possibility that chemicals that have not been measured may also influence the associations presented here.

Further work is needed to explore the effect of individual congeners on breast cancer risk. The use of ridge estimators appeared to offer a useful contribution to our understanding of the effects for these congeners in terms of the classification provided by Wolff *et al.*⁴ However, determination of the ridge coefficient required a certain level of judgement on which different investigators could honestly disagree. It would be important to see whether other studies that measured exposure to individual congeners in adipose tissue arrived at similar conclusions regarding a potential protective effect of congener 156 and adverse effects of congeners 180 and 183.

Acknowledgements

This study is supported in part by a grant #CA-62986 from the National Cancer Institute/National Institute of Environmental Health Sciences. We would like to thank the study participants for their time, co-operation and willingness to contribute samples for chemical analysis. In addition, we appreciate the assistance of the physicians, nurses and related personnel at the Yale-New Haven Hospital. Finally, we are grateful for the careful work of the study interviewers, Donna Carrano, Melita Bosnyak, Heather Hutson, and Sylvia Ullman; the Colorado State University's Analytical Laboratory for their state of the art chemical analyses of adipose tissue contaminants; the Yale Cancer Center Shared Resources for their assistance in the collection of study materials; and Omar Dawood for data management assistance.

References

- McFarland VA, Clarke JU. Environmental occurrence, abundance, and potential toxicity of polychlorinated biphenyl congeners: considerations for a congener-specific analysis. *Environ Health Perspect* 1989;**81**:225-39.
- Safe SH, Zacharewski, T. Organochlorine exposure and risk for breast cancer. Etiology of breast and gynecological cancers: Proceedings of the Ninth International Conference on Carcinogenesis and Risk. *Clin Biol Res* 1997;**396**:133-45.
- Safe SH. Toxicology, structure-function relationship, and human and environmental health impacts of polychlorinated biphenyls: progress and problems. *Environ Health Perspect* 1992;**100**:259-68.

- ⁴ Wolff MS, Camaan D, Gammon M, Stellman SD. Proposed PCB congener groupings for epidemiological studies. *Environ Health Perspect* 1997;**105**:13–14.
- ⁵ Dewailly É, Dodin S, Verreault, R *et al.* High organochlorine body burden in women with estrogen receptor-positive breast cancer. *J Natl Cancer Inst* 1994;**86**:232–34.
- ⁶ Fisher B, Contantino JP, Wickerham DL *et al.* Tamoxifen for prevention of breast cancer: Report of the National Surgical Adjuvant Breast and Bowel Project P-1 Study. *J Natl Cancer Inst* 1998;**90**:1371–88.
- ⁷ Wolff MS, Toniolo PG. Environmental organochlorine exposure as a potential etiologic factor in breast cancer. *Environ Health Perspect* 1995;**103**(Suppl. 7):141–45.
- ⁸ Moysich KB, Ambrosone CB, Vena JE *et al.* Environmental organochlorine exposure and postmenopausal breast cancer risk. *Cancer Epidemiol Biomark* 1998;**7**:181–88.
- ⁹ Zheng T, Holford TR, Tessari J *et al.* Breast cancer risk associated with PCBs by congener. *Am J Epidemiol* 2000;**152**:50–58.
- ¹⁰ Stellman SD, Djordjevic MV, Muscat JE *et al.* Relative abundance of organochlorine pesticides and polychlorinated biphenyls in adipose tissue and serum in women in Long Island, New York. *Cancer Epidemiol Biomark* 1998;**7**:489–96.
- ¹¹ Archibeque-Engle AL, Tessari JD, Winn DT *et al.* Comparison of organochlorine pesticide and polychlorinated biphenyl residues in human breast adipose tissue and serum. *J Toxicol Environ Health* 1997;**52**:285–93.
- ¹² SAS Institute. *SAS/STAT User's Guide. Version 6*. Cary, NC: SAS Institute, 1990.
- ¹³ Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. New York: Chapman & Hall, 1993.
- ¹⁴ Weisberg S. *Applied Linear Regression, 2nd Edn*. New York: John Wiley & Sons, 1985, pp.253–59.
- ¹⁵ Neter J, Wasserman W, Kutner MH. *Applied Linear Regression Models*. Homewood, IL: Richard D Irwin Inc., 1983, pp.393–400.
- ¹⁶ Schaefer RL, Roi LD, Wolfe RA. A ridge logistic estimator. *Commun Statist-Theory Meth* 1984;**13**:99–113.
- ¹⁷ Schaefer RL. Alternative estimators in logistic regression when the data are collinear. *J Statist Comp Sim* 1986;**25**:75–91.
- ¹⁸ Segerstedt B. On ordinary ridge regression in generalized linear models. *Commun Statist-Theory Meth* 1992;**21**:2227–46.
- ¹⁹ Zheng T, Holford TR, Mayne ST, *et al.* DDE and DDT in breast adipose tissue and risk of female breast cancer. *Am J Epidemiol* 1999;**150**:453–58.
- ²⁰ Ambrosone CB, Freudenheim JL, Graham S *et al.* Cytochrome P4501A1 and glutathione S-transferase (M1) genetic polymorphisms and postmenopausal breast cancer risk. *Cancer Res* 1995;**55**:3483–85.
- ²¹ Ishibe N, Hankinson SE, Colditz GA *et al.* Cigarette smoking, cytochrome P450 1A1 polymorphisms, and breast cancer risk in the Nurses' Health Study. *Cancer Res* 1998;**58**:667–17.
- ²² Bailey LR, Roodi N, Verrier CS, Yee CJ, Dupont WD, Parl FF. Breast cancer and CYP1A1, BSTM1, and BSTM1 polymorphisms: evidence of a lack of association in Caucasians and African Americans. *Cancer Res* 1998;**58**:65–70.
- ²³ Matthews HB. Metabolism of PCBs in mammals: routes of entry, storage, and excretion. In: E'rti FM, Kamrin MA (eds). *PCBs: Human and Environmental Hazards*. Boston: Butterworth Publishers, 1983, pp.203–13.
- ²⁴ Sipes IG, Schnellmann RG. Biotransformation of PCBs: metabolic pathways and mechanisms. In: Safe S, Hutzinger O (eds). *Environmental Toxin Series 1*. Berlin: Springer-Verlag, 1987, pp.97–110.
- ²⁵ Wolff MS, Zeleniuch-Jacquotte, A, Dubin N, Toniolo P. Risk of breast cancer and organochlorine exposure. *Epidemiol Biomark* 2000;**9**:271–77.
- ²⁶ Tessari JD, Archibeque-Engle SL. Correspondence re: SD Stellman *et al.*, Relative abundance of organochlorine pesticides and polychlorinated biphenyls in adipose tissue and serum of women in Long Island, New York. *Cancer Epidemiol Biomark* 1999;**8**:111–14.
- ²⁷ Nelder JA, Wedderburn RWM. Generalized linear models. *J R Statist Soc B* 1972;**42**:109–42.

Appendix

The ridge regression estimators for the linear logistic model employed the approach described by Segerstedt¹⁸ and Schaefer,^{16,17} which follows directly from the reweighted least-squares algorithm of Nelder and Wedderburn.²⁷ If n is the number of subjects and p the number of regressor variables, let \mathbf{X} represent an $n \times p$ matrix of regressors. The weights for each observation are derived from the fitted results from the linear logistic model, and these are presented in a diagonal matrix, \mathbf{W} . If \mathbf{Y} is a 0,1 response indicating disease status, and $\boldsymbol{\mu}$ is the expectation, then the Nelder-Wedderburn algorithm makes use of a transformed response vector,

$$\mathbf{Z} = \mathbf{X}\hat{\mathbf{b}} + (\mathbf{Y} - \hat{\mathbf{m}}) \frac{\partial \eta}{\partial \mathbf{m}}$$

where η is the linear predictor, $\log[m/(1 - m)]$. The approximate ridge estimator is given by $\tilde{\mathbf{b}} = (\mathbf{X}'\mathbf{W}\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\mathbf{W}\mathbf{Z}$, where $k(>0)$ is the ridge coefficient and \mathbf{I} an identity matrix.

The variance of the ridge estimator is given by

$$\text{Var}(\tilde{\mathbf{b}}) = (\mathbf{X}'\mathbf{W}\mathbf{X} + k\mathbf{I})^{-1}(\mathbf{X}'\mathbf{W}\mathbf{X})(\mathbf{X}'\mathbf{W}\mathbf{X} + k\mathbf{I})^{-1}$$

In this analysis we were primarily concerned with the collinearity among the congeners, but it was also necessary to adjust for covariates, \mathbf{U} . To accomplish this we used residuals from a prediction of the congener levels from the covariates as the regressors in the formula given above, $\mathbf{X}^* = [\mathbf{I} - \mathbf{U}(\mathbf{U}'\mathbf{U})^{-1}\mathbf{U}']\mathbf{X}$, along with a similar term for the response, $\mathbf{Z}^* = [\mathbf{I} - \mathbf{U}(\mathbf{U}'\mathbf{U})^{-1}\mathbf{U}']\mathbf{Z}$. When used in the expression for the ridge regression estimator given above, this provided values that were adjusted for covariates. The ridge estimators were obtained by using output files from PROC GENMOD in SAS to obtain the fitted values and the weights, which were in turn input into PROC IML for the final step in the calculations.